ANNUAL REPORT

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (OAR)

GRANT [redacted]-AFOSR 69-1803

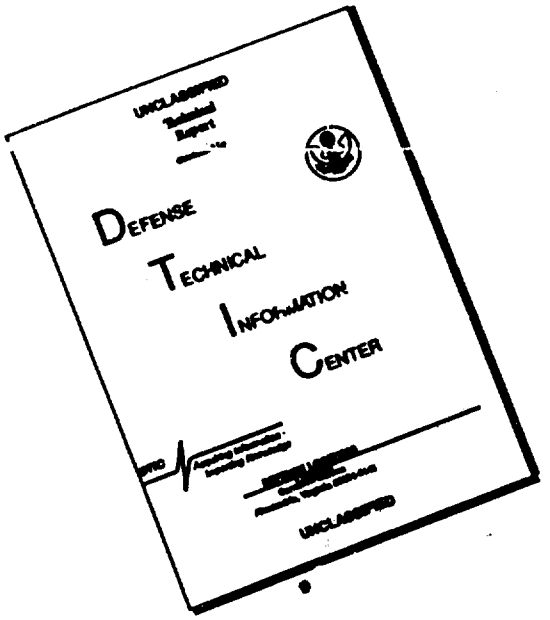
STATISTICAL MODELS

JUNE 1, 1970 - MAY 31, 1971

GEORGE E. P. BOX
Principal Investigator

DEPARTMENT OF STATISTICS
University of Wisconsin
Madison, Wisconsin

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Security Classification

DOCUMENT CONTROL DATA – R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

University of Wisconsin
Department of Statistics
Madison, Wisconsin 53706

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

STATISTICAL MODELS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Scientific Interim

5. AUTHOR(S) *(First name, middle initial, last name)*

George E. P. Box

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 1971	5	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF-AFOSR 69-1803	
b. PROJECT NO. 9749	
c. 61102F	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d. 681304	AFOSR - TR - 71-2692

10. DISTRIBUTION STATEMENT

A. Approved for public release: distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Office of Scientific Research (NAM) 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

Efficient planning requires efficient methods for forecasting and control. One of the objects of the present research is to further extend methods described in a recent successful 550 page book by Box and Jenkins developed under AFOSR sponsorship. Non-stationary models which can adequately represent multiple dependent records developing in time have been obtained and efficient methods for identification, estimation and diagnostic checking have been studied. Of particular importance are canonical forms of the model whereby the information contained in many records can often be summarized in a few composite series. Difficult problems in estimation are being approached using Bayesian methods. In this connection a book on Bayesian inference (by G.E.P.Box and G.C. Tiao) containing some 800 manuscript pages and incorporating research performed under this contract was completed in May 1971 and is being published by Addison-Wesley.

Two problems occurring in continuous time control theory were studied and results obtained on the accuracy properties of numerical algorithms for solving them approximately. Some new results in Bayesian Tolerance Regions have been obtained. In addition to the Bayes book during the reporting period two papers have been published, two have been accepted for publication and five more submitted.

DD FORM 1 NOV 65 1473

Security Classification

Activities and results achieved

1) Time Series Forecasting

The results of many years of previous research sponsored by the Air Force Office of Scientific Research were published in September 1970 in the 550 page book Time Series Analysis Forecasting and Control (G. E. P. Box and G. M. Jenkins published by Holden-Day, San Francisco). One major thrust of the current research is to further extend this work to multiple input-output systems.

A very important class of non-stationary models discussed in the above text and having application in forecasting, scheduling and inventory control consists of the integrated moving average models. These models lead to exponentially weighted averages as best forecasts. Frequently, not one, but many non-stationary series are available each of which can potentially supply information and improve forecasts of the others. In new research, a general class of non-stationary multivariate time series models has been developed which generalizes the integrated moving average models. The model building problem encountered in relating these to data is not easy because it involves the simultaneous estimation of many parameters which appear non-linearly in the model. The approach we are taking involves the investigation of canonical forms. For example, economic indicators and other time series are often closely related and the information one contains is frequently to some extent included in the others. In using our analysis it may be possible to summarize the principal information contained in ten series into say two composite series. When this is possible, problems of forecasting and control are rendered much more tractable.

2) Bayesian Developments

The problems of estimation mentioned above are being tackled by the Bayesian route. One difficulty involved in this approach is the choice of "non-informative" prior distribution appropriate when no previous information exists about the parameters. The problem of choice of non-informative prior distribution has been under extensive study and new results have been included in the book Bayesian Inference by G. E. P. Box and G. C. Tiao the manuscript for which is now complete and was accepted in May 1971 for publication by Addison-Wesley. It contains a great deal of research work sponsored by Air Force Office of Scientific Research to who due acknowledgment is made.

3) Continuous Time Control

The use of Reproducing Kernel Hilbert Spaces (RKHS) as a tool to solve optimization problems occurring in control theory is being studied intensively. The use of this tool is appropriate in problems where time should be considered as a continuous variable, or, where the optimum sampling interval (in time) is to be determined.

i) Techniques involving RKHS were used to develop and study the accuracy properties of a class of computer algorithms for solving linear operator equations numerically. (Technical Report No. 270).

ii) A typical continuous time control problem involves minimizing a quadratic functional subject to a continuous family of linear inequality constraints. The convergence properties of a numerical technique proposed by Daniel for solving this type of problem approximately have been established, using RKHS theory (Technical Report No. 282.)

Both of these problems are important in continuous time forecasting and control.

4) Bayesian Tolerance Regions

The theory of tolerance regions is of considerable practical importance and application. For example, the question as to whether 100β% of rations supplied under a particular distribution policy fullfills minimum maintenence requirements can be answered by constructing a 100β% content statistical tolerance region. There are many other esamples. A Bayesian approach makes it possible to incorporate prior knowledge into the colculations. The necessary distribution theory involves a disguised Wishart distribution which has been recently studied.

5) Colloquim on Statistical Model Building Prediction and Control

A colloquim on Statistical Model Building Prediction and Control was held at the Institute of Electrical Engineers headquarters in Savoy Place, London on 24th April 1971. In a paper by D. J. Reid of the London School of Economics entitled, "A Survey of Statistical Forecasting" the author compared the forecasting ability of various theories when applied to 113 actually occuring time series, consisting of annual quarterly and monthly sample data. The data represented macro-economic variables, commodity prices, and industry wide level of aggregation. Reid found that "the Box-Jenkins method outperformed any of the others by a factor of 3 to 1." These methods were developed under Air Force Office of Scientific Research sponsorship.

6) Meeting of the Royal Statistical Society on Forecasting

A general meeting of the Royal Statistical Society took place in London on January 20, 1971 at which the paper "Dynamic Equations for Economic Forecasting" by J. Bray was read and extensively discussed. Dr. Bray and other speakers emphasized the importance of the forecasting methods developed by Box and Jenkins (under AFOSR sponsorship). G. E. P. Box attended the meeting and contributed to the discussion which is to be published.

Special Honors

Dr. Box was appointed to a distinguished professorship at the University of Wisconsin and now holds the Ronald Aylmer Fisher Chair of Statistics.

Personnel Supported

Principal Investigator -	G. E. P. Box	Summer (Ju.-Aug) 1970
Research Assistants -	C. K. Yang	August 1970
	N. Bhalerao	Acad. Yr. 1970-71
	W. Wei	Acad. Yr. 1970-71
Technical Typist -	C. Smith	June-Aug,1970
		Mar-June 1971

Published

Tan, W. Y. & Cuttman, I., Disguised Wishart variable and a related theorem. Journal of the Royal Statistical Society Series B, Vol. 33 (1971)

Box, G. E. P. & Pierce, D. A., Distribution of residual autocorrelations in autoregressive integrated moving average time series models. Journal of the American Statistical Association, Vol. 65 (1970)

Accepted for publication

Box, G. E. P. and G. C. Tiao, Bayesian Inference in Statistical Analysis. Book of 800 manuscript pages to be published by Addison-Wesley.

Box, G. E. P. and P. Newbold, "Comments on a paper of Coen, Gomme and Kendall" (Journal of the Royal Statistical Society, Series B).

Wahba, Grace, "On the numerical solution of Fredholm integral equations of the first kind," (J. Approx. Theory.)

Submitted for publication

Box, G. E. P., Jenkins, Gwilym, and Guttman, Irwin, "Partial auto-correlations from a Bayesian viewpoint and orthogonal parameterization" (Jour. Amer. Stat. Assoc.)

Guttman, I. & Tan, W. Y., "The use of the disguised Wishart distribution in a Bayesian approach to tolerance region construction. (Ann. Inst. Math. Stat.)

Guttman, I., Pereyra, V., and Scolnik, H.D., "Least squares estimation for a class of non-linear models." (Technometrics)

Wahba, Grace, "A Class of approximate solutions to linear operations equations," (J. Approx. Theory)

Wahba, Grace, "On the minimization of a quadratic functional subject to a continuous family of linear inequality constraints." (SIAM Journal of Control)

Practical Importance of the Work to the Air Force

Efficient forecasting and control methods depend on the building of adequate statistical (probability) models which can properly represent multiple dependent records which are developing in time. The present research is directed toward the building of these models. In this research theory is continually tested and directed by the analysis of the real data.